BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 2: e25637
doi: 10.3897/biss.2.25637

OPEN ACCESS

Conference Abstract

Deep learning for plant identification: how the web can compete with human experts

Hervé Goëau[‡,§], Alexis Joly[|], Pierre Bonnet[§,‡], Mario Lasseck[¶], Milan Šulc[#], Siang Thye Hang[¤]

‡ AMAP, Univ Montpellier, CIRAD, CNRS, INRA, IRD, Montpellier, France
§ UMR AMAP, CIRAD, Montpellier, France
| Inria ZENITH team, Montpellier, France
¶ Museum fuer Naturkunde Berlin, Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany
# Czech Technical University, Prague, Czech Republic
¤ Toyohashi University of Technology, Toyohashi, Japan

Corresponding author: Hervé Goëau (herve.goeau@cirad.fr), Alexis Joly (alexis.joly@inria.fr), Pierre Bonnet (pierre.bonnet@cirad.fr), Mario Lasseck (mario.lasseck@mfn-berlin.de), Milan Šulc (sulcmila@fel.cvut.cz), Siang Thye Hang (hang@kde.cs.tut.ac.jp)

Received: 09 Apr 2018 | Published: 22 May 2018

Citation: Goëau H, Joly A, Bonnet P, Lasseck M, Šulc M, Hang S (2018) **Deep learning for plant identification: how the web can compete with human experts**. Biodiversity Information Science and Standards 2: e25637. https://doi.org/10.3897/biss.2.25637

Abstract

Automated identification of plants and animals has improved considerably in the last few years, in particular thanks to the recent advances in deep learning. In order to evaluate the performance of automated plant identification technologies in a sustainable and repeatable way, a dedicated system-oriented benchmark was setup in 2011 in the context of ImageCLEF (Goëau et al. 2011). Each year, since that time, several research groups participated in this large collaborative evaluation by benchmarking their image-based plant identification systems. In 2014, the LifeCLEF research platform (Joly et al. 2014) was created in the continuity of this effort so as to enlarge the evaluated challenges by considering birds and fishes in addition to plants, and audio and video contents in addition to images.

The 2017-th edition of the LifeCLEF plant identification challenge (Joly et al. 2017) is an important milestone towards automated plant identification systems working at the scale of continental floras with 10.000 plant species living mainly in Europe and North America

© Goëau H et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

illustrated by a total of 1.1M images. Nowadays, such ambitious systems are enabled thanks to the conjunction of the dazzling recent progress in image classification with deep learning and several outstanding international initiatives, aggregating the visual knowledge on plant species coming from the main national botanical institutes. The PlantCLEF plant challenge that we propose to present at this workshop aimed at evaluating to what extent a large noisy training dataset collected through the web (then containing a lot of labelling errors) can compete with a smaller but trusted training dataset checked by experts. To fairly compare both training strategies, the test dataset was created from a third data source, the Pl@ntNet (Joly et al. 2015) mobile application that collects millions of plant image queries all over the world.

Due to the good results obtained at the 2017-th edition of the LifeCLEF plant identification challenge, the next big question is how far such automated systems are from the human expertise. Indeed, even the best experts are sometimes confused and/or disagree with each other when validating images of living organism. A multimedia data actually contains only partial information that is usually not sufficient to determine the right species with certainty. Quantifying this uncertainty and comparing it to the performance of automated systems is of high interest for both computer scientists and expert naturalists. This work reports an experimental study following this idea in the plant domain. In total, 9 deep-learning systems implemented by 3 different research teams were evaluated with regard to 9 expert botanists of the French flora. The main outcome of this work is that the performance of state-of-the-art deep learning models is now close to the most advanced human expertise. This shows that automated plant identification systems are now mature enough for several routine tasks, and can offer very promising tools for autonomous ecological surveillance systems.

Presenting author

Hervé Goëau

References

- Goëau H, Bonnet P, Joly A, Boujemaa N, Barthélémy D, Molino J, Picard M (2011) The ImageCLEF 2011 plant images classication task. In: ImageCLEF (Ed.) Multimedia Retrieval in CLEF. CLEF, Amsterdam, 2011.
- Joly A, Goëau H, Glotin H, Spampinato C, Bonnet P, Vellinga W, Planque R, Rauber A, Fisher R, Müller H (2014) LifeCLEF 2014: Multimedia Life Species Identification Challenges. Lecture Notes in Computer Science. https://doi.org/10.1007/978-3-319-11382-1_20
- Joly A, Bonnet P, Goëau H, Barbe J, Selmi S, Champ J, Dufour-Kowalski S, Affouard A, Carré J, Molino J, Boujemaa N, Barthélémy D (2015) A look inside the Pl@ntNet experience. Multimedia Systems 22 (6): 751-766. https://doi.org/10.1007/s00530-015-0462-9

- Joly A, Goëau H, Glotin H, Spampinato C, Bonnet P, Vellinga W, Lombardo J, Planqué R, Palazzo S, Müller H (2017) LifeCLEF 2017 Lab Overview: Multimedia Species Identification Challenges. Lecture Notes in Computer Science. https://doi.org/10.1007/978-3-319-65813-1_24